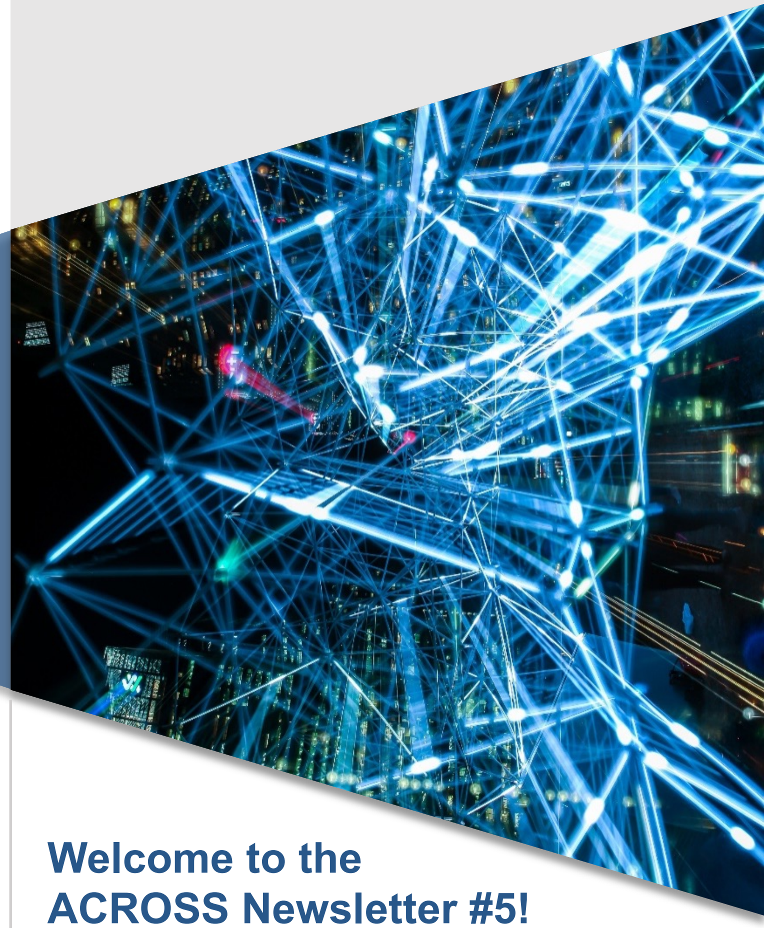
NEWSLETTER #5

SEPTEMBER 2023

ACROSS

HPC Big DAta ArtifiCial intelligence cross Stack PlatfoRm TOwardS ExaScale



INSIDE THIS ISSUE

Welcome to the ACROSS Newsletter #5!

The fifth edition of the ACROSS Newsletter covers one important aspect of any development: validating its behaviour in real contexts.

As such, this issue focuses on the setup of a testbed with the purpose of validating and assessing the capabilities of the ACROSS orchestration stack. To this end, a full-fledged small test infrastructure has been set up upon CINECA G100 Cloud-based resources.

The validation is going to take place during this period and will be based on application workflows derived from Pilot use cases.

Want to know more? Keep up to date on all project news @across_project and sign up for the Newsletter so you never miss out.

Questions? Contact info@acrossproject.eu

This project has received funding from the European High-Performance Computing Joint Undertaking (JU) under grant agreement No 955648. The JU receives support from the European Union's Horizon 2020 research and innovation programme and Italy, France, Czech Republic, United Kingdom, Greece, Netherlands, Germany, Norway.

Orchestrator Testbed

The ACROSS orchestration stack is a complex system integrating diverse software components, each with a specific purpose and providing specific features. Together, these components allow the ACROSS orchestrator to allocate HPC (and Cloud) resources, scheduling the execution of workflow's application steps on top of them, this all by addressing a more *deterministic execution* when compared to the traditional submission of jobs to the batch scheduler.
Validating the capability of this software stack is of primary importance to ensure the proper execution of workflows in production environments. As such, we are planning an extensive experimental campaign, where workflows derived from Pilot use cases and properly scaled down will be used.

Workflow(s) target

ACROSS project aims at providing a flexible, deterministic, efficient way of executing modern workflows composed of steps belonging to diverse domains (i.e., numerical simulations, machine learning model training and inference, HPDA, etc.). To assess the capability of the orchestration stack to support their execution, we target to perform a large experimental campaign.

The first workflow being part of this experimental campaign is derived directly from the ACROSS aeronautic use case referred to as Turbine use case (WP5 – see Figure 1). In this case, the workflow is composed by numerical simulations (LES), and a high-performance data analytics (HPDA) procedure.

To reduce the timeframe for executing an experiment, the entire workflow has been down-scaled, and the URANS (high-fidelity simulations) have been removed, this all without impacting on the complexity of the workflow in terms of step heterogeneity.

Figure 2 provides a graphical representation of the steps involved in the workflow. The last period of the project will see the number of workflows increasing, thus further demonstrating the benefit of the developed orchestration stack.

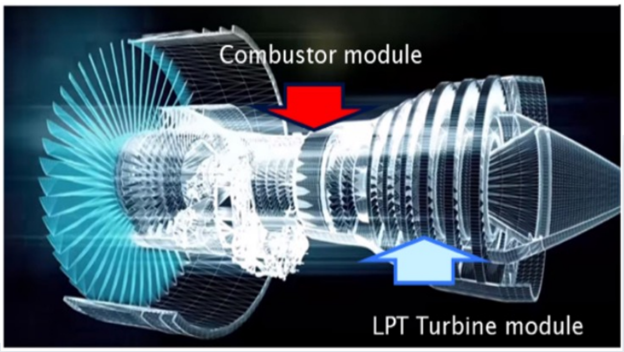


Fig. 1 – ACROSS Turbine use case

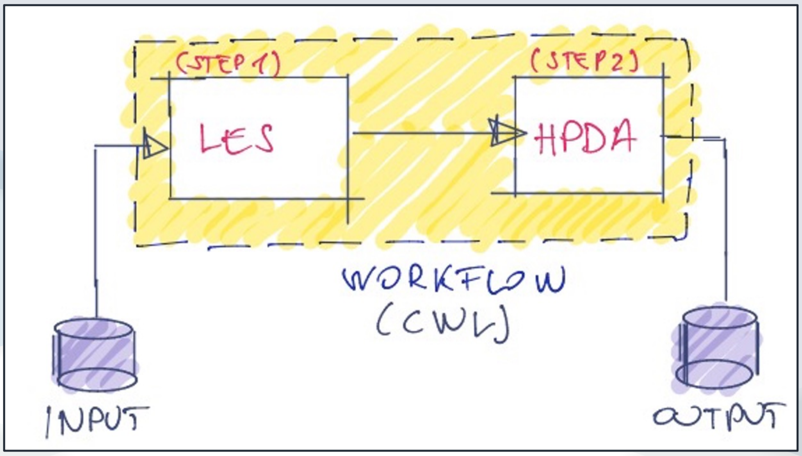


Fig. 2 – Graphical representation of the down-scaled Turbine workflow

CINECA Cloud Infrastructure

Supercomputers are becoming more heterogeneous and modular in order to accommodate different users' needs (AI modelling, data analytics, etc.). As such, resource provisioning models are evolving beyond the queuing system, thus including today Cloud computing models. CINECA is following this approach, by supporting Cloud provisioning model through its ADA Cloud system.

Cloud service and HPC resources

The following points apply (see also Figure 3):

- HPC and CLOUD are crossing paths.
- HPC centres are aware of the change in the user requests and provide more "cloudy" services (https://fenix-ri.eu/).
- 1 Major hyperscalers (Azure, Google, etc.) are expanding their offer to provide HPC services based on high-end technology (e.g., NVIDIA GPUs) and their own HPC hardware (e.g., AWS Graviton, Google TPU).

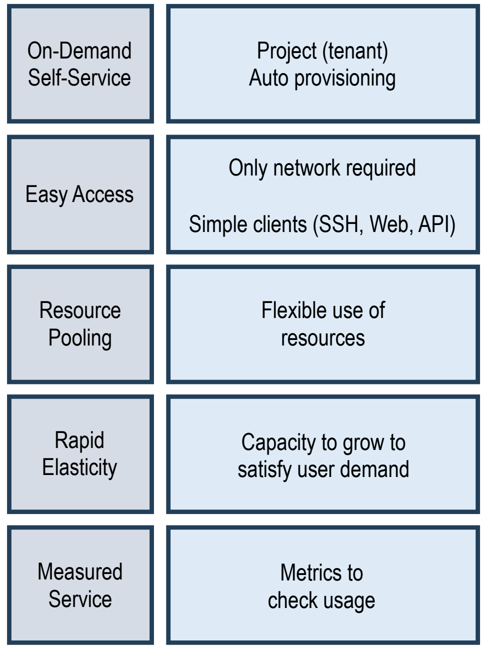


Fig. 3 – Cloud resource provisioning features

Cloud-HPC benefits

User requested features	Motivation	Implementation
Agility	Quickly deploy IT projects	Composable services are provided through OpenStack
Reliability	Disaster Recovery Continuity	DR is not offered. Physical infrastructure redundancy
Scalability and Elasticity	Meet user requirements	Constant growth in capacity
Performance	HPC like SLAs	Offered Not offered, but HD available on working days
Ease of maintenance	Infrastructure managing	Clear division of roles (users/admins)
Security and compliance	GDPR ISO27001	Offered
User Data	Ownership	Users remain owners of their data

CINECA *ADA Cloud*

The HPC cloud infrastructure, named *ADA Cloud* is based on OpenStack Wallaby. It provides:

- 71 interactive OpenStack nodes each equipped with 2 x CPU Intel Cascade Lake 8260, 24 cores each, clocked at 2.4 GHz, and sporting 768 GB of RAM and 2TB of SSD storage.
- 1 PB of *Ceph* dedicated raw storage (i.e., fully based on NVMe/SSD).

This cloud infrastructure is tightly connected both to the LUSTRE storage of 20 PB raw capacity, and to the GSS storage of 6 PB seen by all other infrastructural resources. This setup enables the use of all available HPC systems (i.e., Tier-0 Marconi, Tier-1 Galileo100), addressing HPC workloads in conjunction with cloud resources. It is worth remarking the following points:

- OpenStack Horizon dashboard: https://adacloud.hpc.cineca.it
- User authentication via CINECA IdP, based on Keycloak OpenID. Additional Fenix IdP is going to be available.
- User Guide: https://wiki.u-gov.it/confluence/display/SCAIUS/UG3.5%3A+ADA+Cloud+UserGuide

Multiple cloud images are provided, but users can use their own:

- Multiple cloud flavours are provided, ranging from 1 up to 96 vCPUs, to allocate the entire compute node capability with a single VM. This is to address computing power demanding user workflows.

ACROSS Virtual Cluster

On top of the (Galileo-100) ADA Cloud, a small-scale virtual cluster has been created. The cluster sports a moderate number of (virtual) resources which are backed by high-performance physical ones. The total amount of compute cores (vCPUs) is 288, backed with a total storage volume of 3.0 TiB, and connected with a (high-performance) networking using a flatten topology. For the purposes of the validation process, these resources have been arranged as 19 VMs as follows (see also Figure 4):

- 1 VM equipped with 16 vCPUs used to export the 3.0 TiB of storage as a NFS shared volume.
- 2 VMs equipped with 8 vCPUs each hosting the SLURM batch scheduler, along with other services and orchestration modules (i.e., WARP, StreamFlow, HyperQueue).
- 16 VMs equipped with 16 vCPUs each, used as compute nodes.

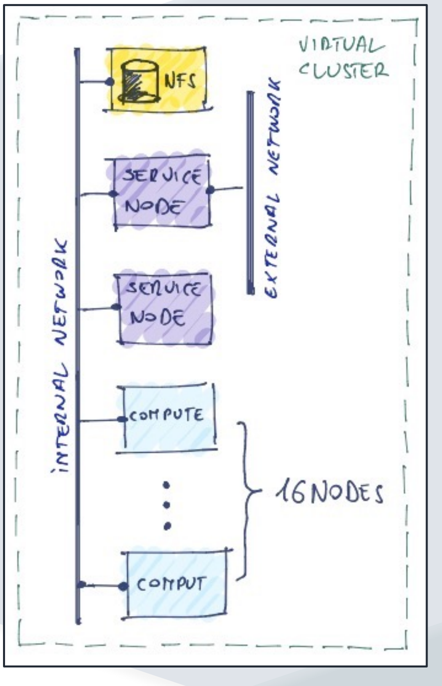


Fig. 4 – ACROSS Virtual Cluster architecture

ACROSS orchestrator validation

An extensive campaign of experiments aimed at comparing the execution of workflows with and without the ACROSS orchestration stack has been planned.

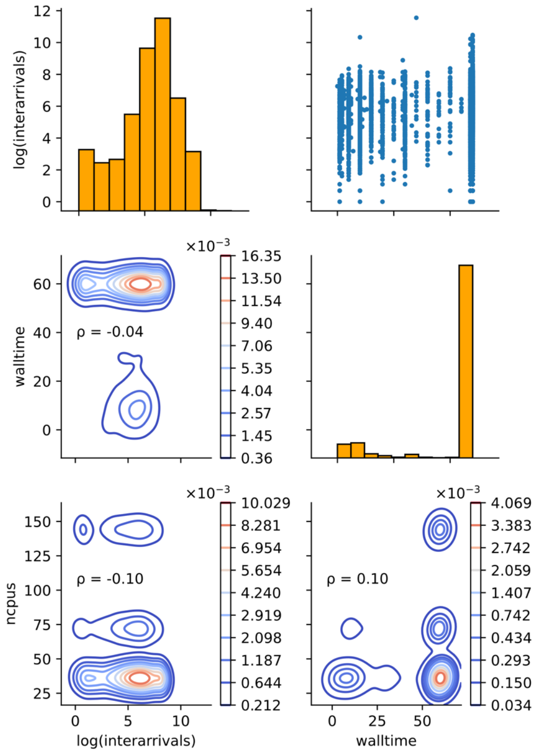


Fig. 5 – Jobs' distribution model

Loading the virtual cluster

One important aspect of performing orchestration validation through an experimental campaign concerns the creation of an 'initial load' on the virtual cluster, such that it reproduces the load condition of a real production system within a controlled environment. To this end, we used historical data collected on a real cluster to derive a statistical model of the jobs submitted. In particular, from the available data, specific features of the submitted jobs have been extracted and used to fit a multi-variated distribution of the jobs over time. This distribution is then used to randomly picking up job's features, create a job with those features and submit it to the system. Once a certain number of jobs have entered the virtual cluster, validation workflow is executed, and the system performance/behaviour is evaluated. Figure 5 provides some graphical information of the jobs' distribution model.

A first glance on experimental results

The workflow (it is meant to be the first one of a set of workflows derived from pilot use cases) as described on page 2 has been implemented using StreamFlow/WARP description format (CWL). A first set of runs have been performed in a clean environment (not loaded with jobs) in order to ensure the correctness of the implementation (see Figure 6). During the next months, a large experimental campaign will be performed to assess the benefit provided by the ACROSS orchestration stack to have a 'deterministic' workflow execution.

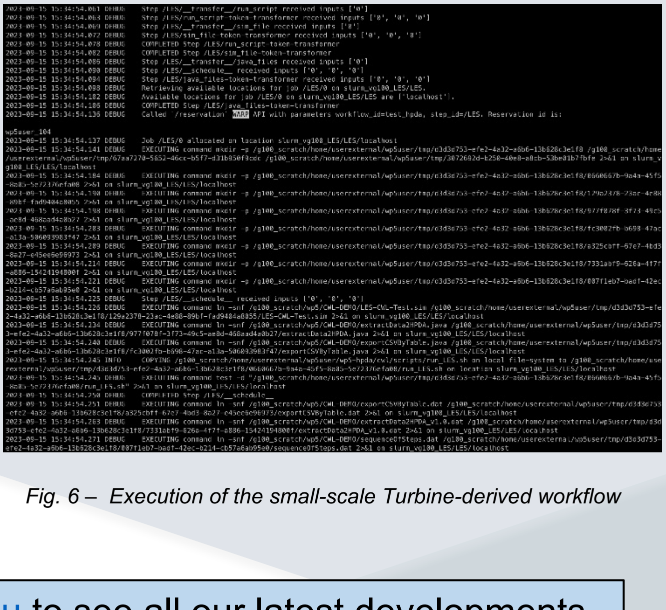


Fig. 6 – Execution of the small-scale Turbine-derived workflow

Visit our website www.acrossproject.eu to see all our latest developments